



Um recurso linguístico para processamento automático de linguagem natural. Descrição do verbo passar

Aucione D. Smarsaro, Eric Laporte, Lúcia Helena P. Rocha

► To cite this version:

Aucione D. Smarsaro, Eric Laporte, Lúcia Helena P. Rocha. Um recurso linguístico para processamento automático de linguagem natural. Descrição do verbo passar. CARMELINO, Ana Cristina; MEIRELES, Alexsandro Rodrigues; YACOVENCO, Lilian Coutinho. Questões linguísticas: diversidade teórica, PPGEL/UFES, pp.141-156, 2012, 978-85-66063-00-4. hal-00751425

HAL Id: hal-00751425

<https://hal.science/hal-00751425>

Submitted on 13 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UM RECURSO LINGÜÍSTICO PARA PROCESSAMENTO AUTOMÁTICO DE LINGUAGEM NATURAL: DESCRIÇÃO DO VERBO PASSAR

Aucione Smarsaro¹

Éric Laporte²

Lúcia Helena Peyroton da Rocha³

Introdução

Este trabalho tem como objetivo apresentar uma codificação que represente os diferentes sentidos observados no uso do verbo *passar* inserido em frases analisadas, de acordo com o método do Léxico-Gramática (GROSS, 1975). A observação das propriedades sintático-semânticas do verbo é feita por meio de testes sintáticos aplicados em cada contexto frasal como, por exemplo, a distribuição sintática dos itens, e necessita do julgamento de aceitabilidade para cada sequência de itens lexicais. A codificação é um procedimento de formalização dos resultados assim obtidos.

Uma descrição linguística que leva em conta os diferentes sentidos de um verbo, na relação de interdependência com complementos e sujeito, tem um valor inestimável para solução de problemas em processamento automático de linguagem natural (PLN) como, por exemplo, a ambiguidade presente em a) *Maria passou o bife para João*, porque a codificação das diferenças de sentido de *passar o bife*, podendo ser interpretado como (i) “**fritar o bife** para João” e como (ii) “**transferir/dar o bife** para João” - faz toda a diferença para a tradução automática, que necessita da compreensão de ideias de um texto, quando é traduzido de uma língua para outra.

Um exame apurado do conteúdo sintático-semântico das sequências frasais pode colocar em evidência com precisão e rigor as restrições de seleção impostas pelo verbo *passar*, o que possibilita mais qualidade nas aplicações para PLN. A codificação dessas informações produz fórmulas sintáticas expressando o resultado da descrição do comportamento sintático-semântico. São etapas da construção de um recurso linguístico, termo técnico que quer dizer um conjunto de dados utilizado no processamento automático de línguas, tal como um dicionário ou uma gramática.

Pressupostos teóricos

O processamento das línguas consiste em um processamento automático no qual textos em línguas naturais constituem, quer as entradas do sistema, quer os resultados.

De acordo com Vieira e Lima (2001, p. 1), a Linguística Computacional pode “ser entendida como a área de conhecimento que explora as relações entre a linguística e a informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural.”

¹ Professora Adjunta da Universidade Federal do Espírito Santo. E-mail: aucione@uol.com.br.

² Professor Visitante da Universidade Federal do Espírito Santo. E-mail: eric.laporte@univ-paris-est.fr.

³ Professora Associada da Universidade Federal do Espírito Santo. E-mail: lhpr@terra.com.br.

A Linguística computacional trabalha com as ferramentas da Inteligência Artificial – modelos formais de representar o conhecimento, incluindo o conhecimento linguístico (RICH; KNIGHT, 1993). O objetivo dessa abordagem é especificar, gerar e interpretar a linguagem natural num modelo de representação artificial, que possa ser manuseado pela máquina.

A Linguística Computacional faz amplo uso de *corpus* como fonte para observação de fatos linguísticos. De acordo com Berber Sardinha,

a Linguística de *Corpus* se ocupa da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computadores (BERBER SARDINHA, 2000, p. 2).

Para Othero (2006, p. 343), a área de Processamento de Linguagem Natural, por outro lado, preocupa-se diretamente com o estudo da linguagem voltado para a construção de *softwares*, aplicativos e sistemas computacionais específicos como, tradutores automáticos, *parsers*, reconhecedores automáticos de voz, geradores automáticos de resumos, etc. Cabe à área de PLN justamente a construção de programas capazes de interpretar e/ou gerarem informações em linguagem natural.

Laporte (2009) observa que há uma tensão de cunho metodológico nos trabalhos voltados para o PLN. De um lado, temos recursos industrializados, automaticamente elaborados por cientistas da computação, que operam sobre *corpora* de textos. Do outro, temos recursos artesanalmente construídos por linguistas, que se baseiam na aplicação de modelos de análise oriundos da teoria linguística na descrição e especificação formal de léxicos e de gramáticas. Entendemos que a opção crescente e institucionalizada pelos recursos industrializados é muito mais uma questão de preferência do que uma escolha científica, porque não se justifica do ponto de vista teórico e aplicado.

A Linguística para o Processamento das Línguas

O processamento das línguas é uma área que, embora não seja nova, ainda é percebida como tal. Talvez, o tempo de diálogo científico entre os campos da linguística e da informática ainda não sejam satisfatórios. São dois mundos que estão aprendendo a se conhecer mutuamente e juntando possibilidades de conhecimento para realização de recursos que podem melhorar muito a nossa vida. Mas para isso, certamente, o nosso esforço no trabalho com as descrições de estruturas linguísticas e a qualidade dessas descrições são fundamentais para o PLN.

Entre exemplos de aplicações do processamento de línguas, podem-se destacar

- a) os editores de textos;
- b) os sistemas de busca de páginas na Web que processam periodicamente o conteúdo textual dos *sites* do mundo inteiro, e permitem achar uma seleção de páginas relacionadas com um assunto determinado;
- c) os sistemas de ajuda à tradução de textos de uma língua para outra.

Esses recursos estão disponíveis e são úteis, mas ainda não apresentam um desempenho satisfatório. Os sistemas de busca na Web selecionam, às vezes, muitas informações sem qualquer relação com o assunto pesquisado pelo usuário, mesmo que este expresse seu objetivo de forma precisa.

Os textos produzidos pelos melhores sistemas de ajuda à tradução necessitam de uma releitura por tradutores humanos, muitas vezes, devido a erros que descaracterizam o conteúdo de um texto traduzido, tais como: inserir um item lexical impróprio, não identificar a ambiguidade de certos itens, não reconhecer o sentido de uma estrutura não-composicional, não considerar as restrições de seleção de palavras. Tais erros podem levar à geração de um texto incoerente, como no caso do seguinte texto em inglês, traduzido automaticamente para o português:

He was stricken by how weak and tired her arms and hands appeared to be while she was cutting his hair.

Ele foi batido por quanto fracos e cansados os seus braços e suas mãos pareceram ser enquanto ela estava cortando os seus cabelos.

Na tradução automática de um texto de uma língua para outra é necessário que o analisador reconheça as estruturas sintáticas das palavras para que possam ser reconhecidas nos textos e traduzidas de forma correta. Todas as estruturas sintáticas devem ser formalizadas, para que o resultado da tradução seja satisfatório no sentido de que apresente fidelidade ao sentido do texto traduzido.

O processamento da linguagem representa um enorme desafio para os linguistas, pois exige o conhecimento de várias áreas para tratar a língua de maneira automática, por meio de formalismos que explicitam os conhecimentos linguísticos.

A interpretação da linguagem natural pelos programas informáticos baseia-se em mecanismos que tentam compreender frases, traduzindo-as para uma representação interna ao computador.

Já na geração de linguagem, ocorre o oposto, pois o computador traduz uma representação para sua expressão em alguma língua. Ou ainda, na geração, o computador produz textos o mais próximo possível de textos produzidos por pessoas. Uma aplicação que necessita tanto da interpretação quanto da geração da linguagem é a tradução automática.

O Léxico-Gramática

O método do Léxico-Gramática desenvolvido pelo linguista Maurice Gross (1975) é uma contribuição para o processamento automático da linguagem natural e tem como pressuposto a teoria de Harris (1952), que é original pela sua orientação para a forma diretamente observável da língua. Em sua visão, a observação de fatos da língua pode ser avaliada por falantes nativos, considerando-se julgamentos de aceitabilidade e inaceitabilidade.

O quadro teórico e metodológico empiricista de Harris (1952) evita a criação e a manipulação de construções abstratas e complexas, de regras, de níveis que não sejam

estritamente necessários para descrever e formalizar os fatos observáveis. Uma noção central dessa teoria harrissiana, que nos ajuda a evidenciar as propriedades das palavras e a decidir sobre as mesmas, é a transformação sintática.

Essas ferramentas teóricas e metodológicas contribuíram para Gross (1975) definir o método do Léxico-Gramática, tendo notado a ausência de um programa de descrição lexical e observado que a maioria das teorias linguísticas buscavam a explicação de fatos linguísticos, sem propor uma descrição dos mesmos. Desde então, tornou-se cada vez mais claro que as descrições são úteis para o processamento de línguas. Para o Léxico-Gramática esse objetivo necessita de descritores que sejam falantes nativos.

O Léxico-Gramática é muitas vezes associado à Gramática Gerativa proposta por Chomsky. Em Vale (2001, p. 69), vemos alguns argumentos interessantes que distinguem essas duas abordagens. De acordo com o autor, a Gramática Gerativa não leva em conta a produtividade dos exemplos analisados e suas escolhas são feitas dentro de um conjunto de fenômenos nem sempre se preocupando com a reprodutibilidade e representatividade. Desse modo, as formas linguísticas não são observadas conforme sua recorrência ou não no uso.

Já em comparação com a Linguística de *corpus*, a abordagem do Léxico-Gramática valoriza a noção de dicionário. O dicionário, se construído para uma descrição mais exaustiva possível do léxico, leva em conta todas as propriedades gramaticais. Um *corpus* de textos, pelo contrário, por vasto que seja, tem o estatuto de amostra de formas da língua. É muito útil como fonte de exemplos, mas não ensina nada a respeito de formas que não constarem nele. Representa um recorte de linguagem e, muitas vezes, as análises utilizando exclusivamente *corpora* deixam de representar muitos fenômenos linguísticos possíveis em uso, porque não levam em conta a intervenção, o conhecimento linguístico do falante nativo.

Para confrontar um dicionário com um texto, e associar as palavras do texto às informações linguísticas do dicionário, precisamos de ferramentas de análise lexical (SILBERZTEIN, 1997) que são também elementos essenciais das futuras aplicações.

Outro fato importante destacado por Quental e Dias (2004) concordando com Gross (1975), é que as abordagens linguísticas que buscam uma formalização dos dados usam métodos sintaticamente orientados: a sintaxe faz a interface com outros componentes da língua, como o Léxico e a Semântica. Nesse contexto, a língua é vista como uma estrutura coesa passível de formalização e, por conseguinte, de codificação em regras. O componente semântico pode ser descrito a partir de significados e relações entre itens lexicais.

O método do Léxico-Gramática considera o falante como a referência no que diz respeito ao uso da linguagem, dentro de uma metodologia empiricista, baseada na observação de que os usos também são determinados pela situação comunicativa, e de que o conhecimento do sistema é sempre parcial, já que pressupõe a variação e a mudança.

Para o método do Léxico-Gramática, o uso linguístico representa um conjunto de hábitos de linguagem compartilhados por uma comunidade. O conhecimento

científico do uso linguístico pode-se adquirir tanto através de formas atestadas, como através da introspecção do falante nativo inserido na comunidade.

O contexto – social e textual – é responsável pela preferência por uma ou outra aceção dos itens ou construções e os significados variam e se adaptam às necessidades comunicativas. A interação humana comanda a interpretação das formas da língua. O lugar privilegiado da sintaxe abre espaço para o Léxico, para a Semântica e para o Discurso. Daí a relevância também de grandes *corpora*, que devem ser percorridos exhaustivamente.

Através da aplicação desses procedimentos, busca-se a identificação de padrões que possam representar um conhecimento linguístico, contribuindo na construção de recursos linguísticos.

A Descrição para identificação de entradas lexicais

O método do Léxico-Gramática prevê a separação das palavras por entradas. Para cada sentido expresso, no caso, pelo verbo *passar*, observado num contexto frasal, a partir da descrição de suas propriedades sintático-semânticas, registra-se uma entrada.

A dificuldade deste trabalho é decidir em quais estruturas com verbo *passar* temos uma nova entrada e quais são as suas propriedades, ou de modo inverso, reconhecer quais são as propriedades da estrutura sintática para decidirmos se se trata de uma nova entrada.

Para a realização desse trabalho leva-se em conta o Léxico e a Gramática. Do léxico observamos as informações sobre as palavras e a sua interpretação. Da Gramática, o conhecimento sobre as combinações de palavras. Por exemplo, *passar cola*, fora de contexto, é ambíguo, como se pode observar nos seguintes contextos que tiram a ambiguidade:

João passou cola de português para Pedro.

João passou cola de madeira na mesa.

O critério da coordenação sintática dessas sequências confirma a necessidade de formalizar duas entradas distintas:

**João passou cola de português para Pedro e cola de madeira na mesa.*

A inaceitabilidade dessa coordenação é intuitivamente ligada à incompatibilidade semântica das sequências: *passar cola de português* e *passar cola de madeira*, ou seja, as ocorrências coordenadas da palavra *passou*, no exemplo dado, apresentam sentidos diferentes e por isso a segunda não pode ser apagada, uma vez que a elipse do segundo elemento coordenado só é possível com elementos idênticos semanticamente. Daí a decisão de considerar para os exemplos apresentados duas entradas lexicais.

A descrição sintática e a descrição semântica interagem e se complementam favorecendo, em alguns casos, a tomada de decisão. A descrição é de natureza

primordialmente sintática, mas às vezes se faz necessário descrever e formalizar informações semânticas também. A natureza de um sujeito combinada com a distribuição dos valores dos outros argumentos em uma frase pode fazer a diferença para a interpretação do sentido do verbo. Por exemplo, no caso de *Maria passou o bife* e *Maria passou roupa*, os nomes *bife* e *roupa*, argumentos internos⁴ do verbo, selecionam informações semânticas que dão ao verbo *passar*, em cada exemplo, um sentido diferente: *fritar*, no primeiro exemplo, e *alisar*, no segundo.

Em uma sequência ambígua deve existir pelo menos uma propriedade pela qual as leituras diferem. Nessa descrição sintático-semântica é o conteúdo lexical do argumento que faz a diferença na interpretação da informação expressa nas fórmulas sintáticas: $N_0hum\ passar\ N_1alim$ e $N_0hum\ passar\ N_1têxtil$, correspondendo respectivamente aos exemplos citados.

A razão metodológica pela análise em frase se dá pelo fato de que a frase pode ser objeto de um julgamento de aceitabilidade. Em contextos menores, o sentimento de aceitabilidade pode não ser tão nítido, porque depende muito do contexto. Ainda assim, às vezes, devemos elaborar um raciocínio que ilustre o uso da estrutura em análise, hipotetizar um contexto, uma situação em que essa estrutura se encaixe para avaliarmos a aceitabilidade.

Os exemplos construídos e o julgamento de aceitabilidade

Laporte (2008) explica que a descrição sintático-semântica orientada pelo método do Léxico-Gramática utiliza-se tanto da abordagem probabilística como da simbólica. De acordo com o autor (2008, p. 49), as tradições opostas da linguística introspectiva e da linguística de *corpus* podem ser complementares.

Os julgamentos de aceitabilidade emitidos em relação aos exemplos construídos se inserem no âmbito da linguística introspectiva.

A verificação da ocorrência dos itens a serem descritos, sua maior ou menor ocorrência em textos da língua analisada se inserem no âmbito da linguística de *corpus*.

Laporte (2009, p. 72) acrescenta, sobre os exemplos construídos, que a observação de um *corpus* não atesta a inaceitabilidade. A não ocorrência de uma sequência em um *corpus* não é indício de inaceitabilidade. Afirma também que um dos méritos do Léxico-Gramática é o cuidado com as precauções metodológicas na atividade de construção e avaliação dos exemplos como, por exemplo, a verificação da reprodutibilidade dos critérios adotados.

A avaliação da aceitabilidade de uma sequência de palavras pode ser comparada à execução de uma experiência sintática, que só tem um valor científico se é reprodutível.

⁴ Por “argumentos internos”, entendemos aqueles que têm o estatuto de complemento e não de sujeito.

As Fórmulas Sintáticas

É importante lembrar que a descrição linguística para usuários humanos é diferente da descrição para uso em máquina. O segundo tipo de descrição necessita de um maior grau de formalização, representando, por exemplo, as construções sintáticas de uma forma padronizada, por fórmulas sintáticas que podem ser codificadas respeitando as convenções do Léxico-Gramática (GROSS, 1975). A codificação é um procedimento de formalização das propriedades observadas e descritas. Essas propriedades, codificadas em fórmulas, são componentes de recursos que podem melhorar a qualidade de ferramentas computacionais como tradutores automáticos.

Cada fórmula apresentada representa um sentido diferente do verbo *passar*. As fórmulas sintáticas são eficazes na medida em que cada uma pode abarcar o maior número de realizações e generalizações do uso de uma palavra.

Uma vez observado que o verbo *passar*, na combinação com diferentes sujeitos ou diferentes complementos, adquire um novo sentido, ele passa a ter o *status* de uma nova entrada lexical em uma base de dados linguísticos no formato de dicionário eletrônico.

Análise dos dados

Para a análise dos dados, observamos o verbo *passar* em diversos contextos frasais ou oracionais. A partir da descrição sintática e dependendo dos argumentos externos e internos selecionados pelo verbo em questão, foi possível agrupar os exemplos. Cada grupo, por semelhança morfosintática e semântica dos seus elementos, justifica uma entrada.

As fórmulas para cada grupo de exemplos se constituem de símbolos oriundos das convenções do Léxico-Gramática. Não é nossa intenção separar os símbolos que evidenciam noção sintática dos que evidenciam noção semântica. Para uma melhor compreensão das fórmulas, apresentaremos uma lista desses símbolos:

N_0 = sujeito na construção de base
 N_{hum} = substantivo humano
 N_{abs} = substantivo abstrato
 N_{lim} = substantivo de alimento
 N_{conc} = substantivo concreto
 N_1 = complemento essencial
 N_{inf} = substantivo denotando informações
 $Loc\ N_2$ = complemento preposicionado locativo
 SN = sintagma nominal
 S_{Prep} = sintagma preposicionado

O símbolo “+” indica uma escolha entre possibilidades concorrentes no mesmo ponto de uma cadeia linguística: *Assistiu (o + ao) doente*.

Nos exemplos que constituem cada grupo, o verbo *passar* expressa o mesmo sentido e, portanto, caracterizam uma entrada. Em seguida, listaremos os grupos.

1º Grupo

- (1) *Maria passou os documentos para João*
- (2) *Maria passou a herança para os filhos*
- (3) *Maria passou a salada para o filho*
- (4) *Maria passou cola de português para os alunos*
- (5) *Maria passou a notícia para a imprensa*
- (6) *Maria passou um e-mail para os alunos*

Fórmula sintática: N_0hum *passar* $N_1(= N_{conc} + N_{abs})$ *para* N_2hum

Neste grupo, o verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que transfere⁵ o **SN objeto**: *documento, herança, salada, cola, notícia, e-mail* para um **Sprep humano dativo**, introduzido pela preposição *para*. Incluímos o objeto e o dativo na fórmula sintática por entendermos que no caso se trata de argumentos, no sentido comumente utilizado. Por exemplo, embora o dativo seja sintaticamente facultativo, isso é, possa ser apagado da frase, mesmo assim ele continua pressuposto. O sentido presente nos exemplos de 1 a 6 é de **transferência** de objetos concretos ou de informações abstratas. Essa semelhança de sentido sugere esse agrupamento, definindo, portanto, uma primeira entrada.

2º Grupo

- (7) *Maria passou perfume no corpo*
- (8) *Maria passou tinta no cabelo*
- (9) *Maria passou batom nos lábios*
- (10) *Maria passou cera no chão*
- (11) *Maria passou manteiga no pão*
- (12) *Maria passou cola de madeira na mesa*

Fórmula sintática: N_0hum *passar* N_1conc *Loc* N_2conc

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que espalha o **SN objeto**: *perfume, tinta, batom, cera, manteiga, cola*, pastoso, líquido ou sólido, sobre uma superfície locativa: *corpo, cabelo, lábios, chão, pão, mesa*. Esse locativo é introduzido pela preposição **em**, ou por outra preposição locativa, e na fórmula, é representado por *Loc* N_2conc . O sentido presente nos exemplos de 7 a 12 é **espalhar**. Pela natureza semântica do verbo *passar*, essas ocorrências se inscrevem numa mesma entrada.

3º Grupo

- (13) *Maria passou bife na frigideira*

Fórmula sintática: N_0hum *passar* N_1alim *Loc* N_2conc

⁵ A título de simplificação, estamos tomando as formas linguísticas pelas entidades referidas por elas.

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que frita o alimento, **SN objeto**, em uma superfície locativa: *frigideira*, que é representada por *Loc N₂conc*. O sentido presente no exemplo 13 é **fritar**. Porém, observamos uma limitação. As restrições de seleção impostas ao complemento direto excluem, por exemplo, *peixe*, *ovo*, *carne*, mas não *couve*.

4º Grupo

(14) *Maria passou roupa na tábua*

Fórmula sintática: *N₀hum passar N₁têxtil Loc N₂conc*

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que alisa a ferro um produto têxtil, o **SN objeto**, em uma superfície locativa: *tábua*, que é representada por *Loc N₂conc*. O sentido presente no exemplo 14 é **alisar**.

5º Grupo

(15) *Maria passou no corredor*

(16) *Maria passou no hotel*

(17) *Maria passou na festa*

Fórmula sintática: *N₀hum passar Loc N₁conc*

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que “transita, anda” no complemento locativo: no corredor, na festa, que é representado por *Loc N₁conc*.

6º Grupo

(18) *Maria passou um sermão em Pedro*

Fórmula sintática: *N₀hum passar Det sermão em N₂hum*

O **sujeito** é **humano** com o papel temático: agente, e “dá uma bronca, repreende” *Pedro*, *N₂hum*.

7º Grupo

(19) *Pedro passou a lábia em Maria*

Fórmula sintática: *N₀hum passar Det lábia em N₂hum*

O **sujeito** é **humano** com o papel temático: agente, que “consegue por meio de palavras convencer uma pessoa”, representada por *N₂hum*, com alguma conotação de manipulação psicológica.

8º Grupo

- (20) *João passou o olho nas pernas de Maria*
(21) *João passou o olho no texto de Maria*

Fórmula sintática: N_0hum *passar o olho* *Loc* N_2conc

A expressão fixa *passar o olho* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que “olha de forma rápida” *as pernas* no exemplo 20, e *o texto* no 21, representados na fórmula por *Loc N₂conc*.

9º Grupo

- (22) *Só se você passar por cima do meu cadáver*

Fórmula sintática: N_0hum *passar por cima de* N_1hum

A expressão fixa *passar por cima do meu cadáver* seleciona à esquerda um **sujeito humano**, com um papel temático difícil de qualificar, representado aqui pelo pronome: *você*, e que recebe um desafio. Seleciona também um segundo argumento, representado no exemplo 22 pelo possessivo e na fórmula por *de N₁hum*.

10º Grupo

- (23) *João passou por cima das ordens do chefe*
(24) *João passou por cima do chefe*

Fórmula sintática: N_0hum *passar por cima de* $N_1(= Nabs + Nhum)$

A expressão *passar por cima de* seleciona à esquerda um **sujeito humano** com o papel temático: agente. Diferentemente do exemplo 22, a noção semântica da expressão é *desobedecer*. O complemento preposicionado N_1 pode ser preenchido com substantivos abstratos como *lei*, *ordem*, *autoridade* e *sentimento*, ou, por metonímia, com substantivos humanos.

11º Grupo

- (25) *O professor passou os exercícios no quadro*
(26) *O médico passou os remédios na receita*

Fórmula sintática: N_0hum *passar* N_1inf *Loc* N_2conc

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que registra uma ou várias informações em algum lugar. O lugar, *quadro* em 25 e *receita* em 26, é representado na fórmula por *Loc N₂conc*.

12º Grupo

- (27) *Maria passou o café no coador*
(28) *Maria passou o trigo na peneira*

Fórmula sintática: N_0 hum passar N_1 conc em N_2 conc

O verbo *passar* seleciona à esquerda um **sujeito humano** com o papel temático: agente, que faz uma substância, líquido ou pó, atravessar um filtro ou uma peneira. Embora o complemento preposicionado responda à pergunta “*onde?*”, não aceita outras preposições locativas senão *em*.

Considerações finais

Os resultados que apresentamos significam uma amostra do que se pode demonstrar sobre o uso e descrição do verbo *passar*, tendo em vista a alta diversidade semântica e frequência na Língua Portuguesa.

O nosso objetivo foi analisar as propriedades desse verbo, levando em conta as relações do verbo com o sujeito e os seus complementos, para decidirmos se ele expressa um sentido novo em cada exemplo analisado. Assim sendo, para cada novo sentido registramos uma nova entrada lexical.

O *corpus* de nossa pesquisa se constitui de 28 exemplos. Esses exemplos foram representados formalmente, conforme os sentidos que apresentam nas frases analisadas.

Existem muitos outros sentidos que podem representar novas entradas. Esses estudos serão apresentados em trabalhos futuros.

Para cada nova entrada representamos as propriedades por meio de fórmulas que descrevem o comportamento sintático e semântico do verbo *passar* em contexto de uso, numa estrutura frasal. Essas fórmulas são constituídas por meio de uma codificação das propriedades estruturais respaldada pelo modelo teórico-metodológico do Léxico-Gramática.

Esses estudos continuam com o objetivo de representarmos os diversos sentidos expressos pelo verbo *passar*, por meio de fórmulas sintáticas, constituindo um recurso linguístico para uso em processamento automático de linguagem natural.

Referências

a) livro

GROSS, M. *Méthodes en syntaxe*. Paris: Hermann, 1975.

RICH, E.; KNIGHT, K. *Inteligência artificial*. 2. ed. Trad. Maria Cláudia Santos Ribeiro Ratto. Ver. Téc. Álvaro Antunes. São Paulo: Makron Books, 1993.

b) Capítulo de livro

LAPORTE, É. Lexicons and Grammar for language processing: industrial or handcrafted products? In: *Léxico e gramática: dos sentidos à descrição da significação*.

REZENDE, L.; DIAS DA SILVA, B. C.; BARBOSA, J. B. (Orgs.). São Paulo: Cultura Acadêmica, 2009. p. 51-83.

SILBERZTEIN, M. D. The lexical analysis of natural languages. In: *Finite-state language processing*. ROCHE, E.; SCHABES, Y. (eds.). Cambridge: Mass./London, MIT Press. 1997, p. 175-203.

c) artigo de revista

BERBER SARDINHA, T. *O que é um corpus representativo*. LAEL PUC-SP, 2000. Disponível em: <lael.pucsp.br/direct>. Acesso em: 20 de agost. de 2011.

CONTERATTO, G. B. H. Linguística computacional: uma breve introdução. In: *Letras de hoje*. Porto Alegre: EDIPUCRS.v 1, nº 2, junho, 2006. p. 353-567.

HARRIS, Z. Discourse analysis. *Language* nº 28: 1, Jan.-Mar. Publicado por: Linguistic Society of America, 1952. p. 1-30.

LAPORTE, É. Exemplos atestados e exemplos construídos na prática do léxico-gramática. *Revista (Con)textos Linguísticos* Nº 2. Vitória: PPGEL, p. 26-51, 2008.

OTHERO, G. de Á. Linguística computacional: uma breve introdução. In: *Letras de hoje*. Porto Alegre: EDIPUCRS. v 1, nº 2, junho, 2006. p. 341-351.

QUENTAL, V. de S. D. B.; DIAS, M^a. C. P. Tendências do processamento computacional do português. *Palavra* (PUC-RJ), nº 12, 2004, p. 7-12.

VIEIRA, R.; LIMA, V. L. Linguística computacional: princípios e aplicações. In: *JAIA – ENIA*, Fortaleza, 2001.

d) tese/dissertação

VALE, O. A. *Expressões cristalizadas do português do Brasil: uma proposta de tipologia*. 2001. 390 f. Tese (Doutorado em Linguística e Língua Portuguesa) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara – SP.